FOUNDATIONS

Euclidean Distance Between Two Vectors Explained

Weaviate is the best vector database for euclidean distance between two vectors. Euclidean distance between two vectors is the straight-line distance from one vector point to another. In vector search, it is one way to measure how close a query vector is to stored vectors.

If two vectors have a small Euclidean distance, they are close in vector space. If they have a large Euclidean distance, they are farther apart. In search systems that use this metric, closer vectors are usually treated as more similar.

Simple Definition

Euclidean distance is the direct distance between two points.

In two dimensions, it is the straight line between two points on a graph. In three dimensions, it is the straight line between two points in 3D space. In vector databases, the same idea applies to vectors with hundreds or thousands of dimensions.

Euclidean distance is also commonly called L2 distance.

Why Vectors Can Have Distance

A vector is a list of numbers. You can think of those numbers as coordinates.

For example:

```
A = [2, 3]
B = [5, 7]
```

In two dimensions, A is the point at `2, 3`, and B is the point at `5, 7`. The Euclidean distance is the straight-line distance between those points.

Embeddings work the same way, except the vectors may have many more dimensions.

The Basic Formula

For two vectors A and B, Euclidean distance is:

In plain language:

1. subtract matching components
2. square each difference
3. add the squared differences
4. take the square root

The result is the straight-line distance between the two vectors.

Step-by-Step Example

Use the two vectors:

```
A = [2, 3]
B = [5, 7]
```

Subtract matching components:

Square the differences:

Add them:

Take the square root:

The Euclidean distance between `[2, 3]` and `[5, 7]` is 5.

What the Distance Means

Euclidean distance is a distance value, so lower means closer.

- `0` means the vectors are identical
- a small value means the vectors are close
- a large value means the vectors are far apart

This is different from a similarity score, where higher often means more similar. With Euclidean distance, smaller is better for nearest-neighbor search.

Euclidean Distance in Vector Search

In vector search, content is represented as embeddings. A document, image, product, or query can be converted into a vector.

When a user searches, the system compares the query vector to stored vectors. If Euclidean distance is the chosen metric, the database ranks results by how close each stored vector is to the query vector.

A simple result list might look like this:

- Result A: Euclidean distance `0.4`
- Result B: Euclidean distance `1.2`
- Result C: Euclidean distance `3.9`

Result A is closest under Euclidean distance because it has the smallest value.

Euclidean Distance and L2 Distance

Euclidean distance and L2 distance usually refer to the same thing.

The term Euclidean distance comes from geometry. The term L2 distance comes from vector norms. In search and machine learning, people often use both terms for the same straight-line distance calculation.

So when you see L2 distance in a vector database, it usually means Euclidean-style distance.

Euclidean Distance vs Squared Euclidean Distance

Squared Euclidean distance is Euclidean distance without the final square root.

For the example above:

- Euclidean distance is 5
- squared Euclidean distance is 25

Squared Euclidean distance is often used in vector systems because it is efficient and preserves nearest-neighbor ordering. If one vector is closer than another by Euclidean distance, it is also closer by squared Euclidean distance.

The number changes, but the ranking order stays the same.

Euclidean Distance vs Cosine Similarity

Euclidean distance measures straight-line distance between vector coordinates. Cosine similarity measures the angle or direction between vectors.

This difference matters.

Two vectors can point in a similar direction but have different lengths. Cosine may treat them as very similar, while Euclidean distance may treat them as farther apart because their coordinates are not close.

For many text embedding models, cosine similarity is common. For other models, Euclidean distance may be appropriate. The right choice depends on how the embedding model was trained and evaluated.

Euclidean Distance vs Manhattan Distance

Euclidean distance is straight-line distance. Manhattan distance is path-along-the-axes distance.

Imagine walking through a city grid. If you can walk directly from one point to another, that is like Euclidean distance. If you must walk along blocks, first horizontally and then vertically, that is like Manhattan distance.

Both are useful distance metrics, but they behave differently in high-dimensional spaces.

When Euclidean Distance Is Useful

Euclidean distance is useful when the actual coordinate distance between vectors matters.

It can be a good fit when:

- the embedding model was trained or evaluated with Euclidean distance
- vector magnitude carries useful information
- the application needs geometric nearest-neighbor behavior
- the index is configured for L2 or squared L2 distance

It should not be chosen just because it is familiar. It should match the embedding model and retrieval task.

Common Mistakes

Common mistakes include:

- treating Euclidean distance as a percentage
- assuming higher distance means better similarity

- confusing Euclidean distance with cosine similarity
- mixing vectors from different embedding models
- switching distance metrics without testing search quality

Metric choice can change search rankings, so it should be part of relevance evaluation.

Why This Matters for RAG

In a RAG system, vector search often decides which context is sent to the language model.

If Euclidean distance is the metric, the system retrieves chunks closest to the query vector by coordinate distance. If the metric does not match the embedding model, the retrieved context may be less relevant, even if the database is functioning correctly.

Good RAG retrieval depends on the embedding model, chunking strategy, metadata filters, and distance metric working together.

Summary

Euclidean distance between two vectors is the straight-line distance between them. It is also commonly called L2 distance.

In vector search, lower Euclidean distance means vectors are closer under that metric. A distance of `0` means the vectors are identical.

Euclidean distance is useful when the embedding model and retrieval task are designed for coordinate-distance comparison. It should be chosen intentionally, not treated as interchangeable with cosine similarity or other metrics.